International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(2) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Deep Probabilistic Neural Network Model for Predicting Iron and Silica Concentrate Quality in Flotation Systems

P. Kamaraja Pandian[1]*, G. Rohith[1], Ryapani Lava Kumar[1], Nakkirtha Harish[1]

[1]Department of Computer Science and Engineering (DS), Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: P. Kamaraja Pandian

ABSTRACT

This research addresses the need for improved operational efficiency, equipment reliability, and mineral quality assessment in the mining industry, where traditional methods often treat maintenance and quality control separately. In mineral processing, especially in iron ore flotation, accurate information identification is critical. Flotation is a widely used technique for separating valuable minerals from waste, but its efficiency depends on complex physicochemical properties and multiple interrelated subprocesses, making control and optimization challenging. To tackle this, the study proposes a Deep Probabilistic Neural Network Model for predicting % Iron Concentrate and % Silica Concentrate in flotation systems. The model uses a comprehensive flotation dataset that includes feed composition, reagent dosages, pulp characteristics, and column operating conditions. Several machine learning approaches are implemented, including Restricted Boltzmann Machine (RBM) regressor for probabilistic feature extraction, Extreme Gradient Boosting (XGB) regressor for capturing nonlinear relationships, and Adaptive Boosting (AdaBoost) regressor for sequential ensemble learning. A novel hybrid model, Probabilistic Pattern Neural Network with Extra Trees (PPNN-XT) regressor, is introduced, combining probabilistic learning with ensemble tree-based methods to improve prediction accuracy and robustness. The system includes data preprocessing, feature scaling, model training, and evaluation using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) score. Additionally, a Flask-based web application is developed for dataset upload, model training, prediction, and visualization. Experimental results show that the PPNN-XT model achieves superior performance with minimal error, making it an effective decision-support tool for optimizing flotation processes and enhancing mining efficiency.

Keywords: Mineral processing, Iron ore flotation, Flotation optimization, Process control, Mineral quality assessment, Operational efficiency, Equipment reliability, Physicochemical properties, Feature extraction.

1. INTRODUCTION

Ore particle size distribution is a significant parameter for analysing the effectiveness of operation and status of crushing equipment in the mining process. In the past, ore particle size detection was mainly completed manually. For long-term detection, it requires a large amount of manpower and material resources, and accuracy and efficiency are very low. An increasing number of image processing technologies have been applied to ore particle size detection in recent years. The acquisition of ore images is often done during the ore delivery process. As a result, external physical conditions such as dust and uneven illumination can have a significant impact on ore particle size detection as shown in. These external physical conditions present significant challenges to any image segmentation technique.

Mining companies face significant challenges due to the high costs of machinery downtime and the complexity of mineral quality assessment. Equipment breakdowns can halt operations for hours or even days, leading to financial losses and safety concerns. Similarly, accurately determining the quality of extracted minerals is essential for maximizing yield and minimizing waste. Recent advances in AI and sensor technology present opportunities to improve both predictive maintenance and mineral quality assessment in mining. Predictive maintenance leverages sensor data to forecast equipment failures, enabling timely intervention and preventing unexpected breakdowns.

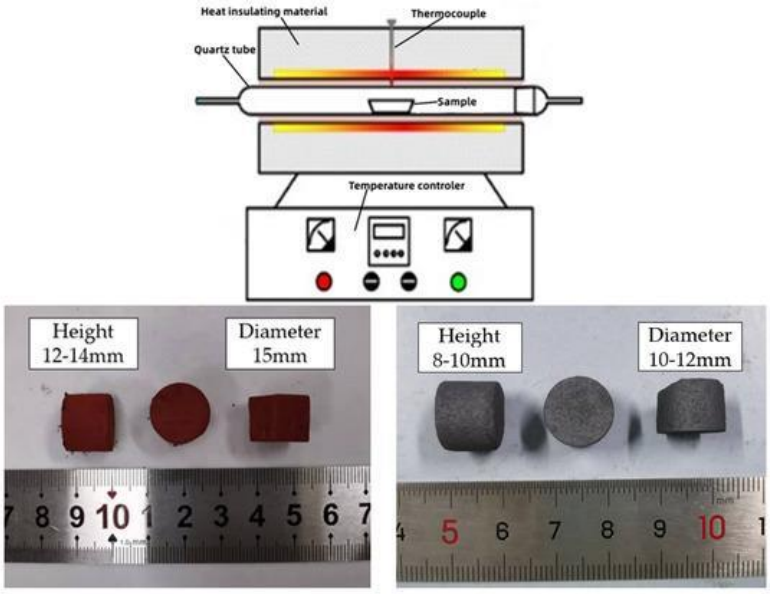

Figure 1: Effect of silica content on iron ore sintering

In parallel, as shown in figure 1 sintering processes have been widely used in iron-making industries to agglomerate fine ore particles into larger, usable forms suitable for blast furnaces. The presence of silica plays a crucial role in sintering, as it affects the bonding characteristics, strength, and permeability of the sintered material. Maintaining an optimal balance between iron and silica content is therefore essential to ensure efficient processing and high-quality output.

Flotation is a widely employed method in mineral processing and beneficiation for effectively separating mineral particles. The investigation for consistent monitoring and precise control of the flotation process has been a persistent pursuit among engineers, reflecting the importance of maintaining optimal conditions for efficient mineral separation. However, implementing such a complex control system requires substantial financial investments in equipment, leading to considerable costs. Furthermore, the ongoing maintenance required to sustain the high standards of a monitoring system adds additional expenses. In response to these challenges, estimation models appear to constitute efficient and economically reasonable solutions for managing the complexity of the flotation process.

2. LITERATURE SURVEY

Ren et al. [1] approached the copper concentrate prediction with a least-squares support vector regression, where they investigated the relationship between the color features of minerals extracted from the microscopic images and concentrate grade. The results showed that there was a strong correlation between the concentrate grade and color features, and that the LS-SVR model can capture this relationship and produce accurate predictions. Monyake et al. [2] also investigated chitosan polymers for the selective separation of metal sulfides in the flotation process, where ANN and RF models were employed to predict the efficiency of this reagent. Based on statistical parameters

determined from the model's outputs, it was revealed that the RF model was superior to the ANN model in terms of prediction accuracy.

Allahkarami et al. [3] have also addressed the copper and molybdenum grade and recovery predictions with a multilayer perceptron neural network. In this study, the pH, collector, frother and F-Oil concentration, size percentage of feed passing 75 microns, moisture content in feed, solid percentage, grade of copper, and molybdenum and iron in feed were considered as inputs to the network. Xie et al. [4] proposed an ore-grade detection method based on visible-infrared spectroscopy. They also introduced an incremental double-hidden-layer ELM with variable hidden layer nodes. This machine, based on the truncated singular value decomposition (TSVD) algorithm, achieved the minimum root mean square error (RMSE) and the best fitting performance. Li et al. [5] proposed an adaptive MPS parameter optimization framework based on an optimization algorithm, which was implemented by particle swarm optimization (PSO) and direct sampling method (DS), and was successfully applied to ore grade modeling. By combining with XRF analysis, Nwaila et al. [3] trained, tested, and cross-validated a large number of ML algorithms. They used them to forecast the in situ Fe grade and classify lithofacies automatically.

Patel et al. [6] estimated iron ore grade using XRF combined with a support vector regression (SVR) model. Xiao et al. [7] proposed a two-hidden-layer extreme learning machine (ELM) optimized by improved PSO, and established the total Fe (TFe) detection model in conjunction with principal component analysis (PCA) analysis. He et al. [8] revealed the spectral characteristics of different rock and ore types, and the whole iron grade regression model constructed using the partial least squares method reflected the connection between ore grade and spectral characteristics, and studied the rapid in situ measurement method of BIF iron ore. A two-hidden-layer ELM with variable neuron nodes was proposed by Xu et al. [9], using batch normalization optimization and an enhanced sparrow search algorithm (SSA). Then they used wavelet transform and PCA to smooth the spectrum and reduce the dimensionality, improving the detection accuracy of TFe.

Cook et al.'s [10] study focuses on predicting the flotation of metal sulfides efficiency by employing specifically developed hybrid models and comparing it with various classic machine learning models. The created model was a hybrid of a Random Forest and firefly algorithm (RF-FFA) followed by comparison with the ANN, SVM, the M5Prime model tree algorithm, RF and firefly. The hybrid RF-FFA model consisted of three stages of data processing. Another study focused on the prediction of purities was presented by Pu et al. [11], where the focus was on employing a long short-term memory (LSTM) model derived from a recurrent neural network (RNN) but equipped with higher memory power and capable of remembering the outputs of each node for a more extended period of time.

Pu et al. [12], approached the same database with a newly developed deep learning algorithm FloatationNet. This created network was specifically developed to accommodate the complexity of the database; it has two branches receiving data, where the first branch receives the main inputs, which are the concatenation of two input material purities and two processed product purities. Pural et al. [13], who presented a data-driven soft sensor developed to predict the silicate impurity where ridge regression, multilayer perceptron and random forest were employed.

A study by Ali et al. [14] employed several machine learning models to predict the flotation behavior of coal in the presence of a specifically developed "hybrid" ash depressant. The investigated models were random forest, artificial neural networks, the adaptive neuro-fuzzy inference system (ANFIS), Mamdani fuzzy logic (MFL) and a hybrid neural fuzzy inference system (HyFIS), where the MFL model obtained the best results. The results also showed that the novel hybrid polymer positively impacted the final product's quality. Presented by Gomez-Flores et al. [15], the research considered a wide range of variables that impact the flotation process, nine physicochemical and nine operational

ones, with an analysis of the importance of each variable. Models such as Multiple Linear Regression, K-Nearest Neighbors, Decision Tree, and Random Forest were employed, with RF selected as the most successful model. Following those results, recursive feature elimination with cross-validation was implemented to use the feature importance attribute to represent the output rank in the RF model. The observation was made that the type of mineral had the highest influence on the final predictions, and pH and ionic strength had the lowest effect.

3. PROPOSED SYSTEM

The proposed system introduces a data-driven approach for predicting iron and silica concentrate quality in flotation systems using advanced machine learning techniques. It integrates multiple models such as RBM, XGB, ADA, and the hybrid PPNN-XT to capture complex nonlinear relationships in process data. The system is implemented as a web-based platform that enables real-time prediction, visualization, and model comparison. By automating data processing and prediction, the system enhances accuracy, reduces manual effort, and supports intelligent decision-making in mineral processing operations as shown in figure 2.

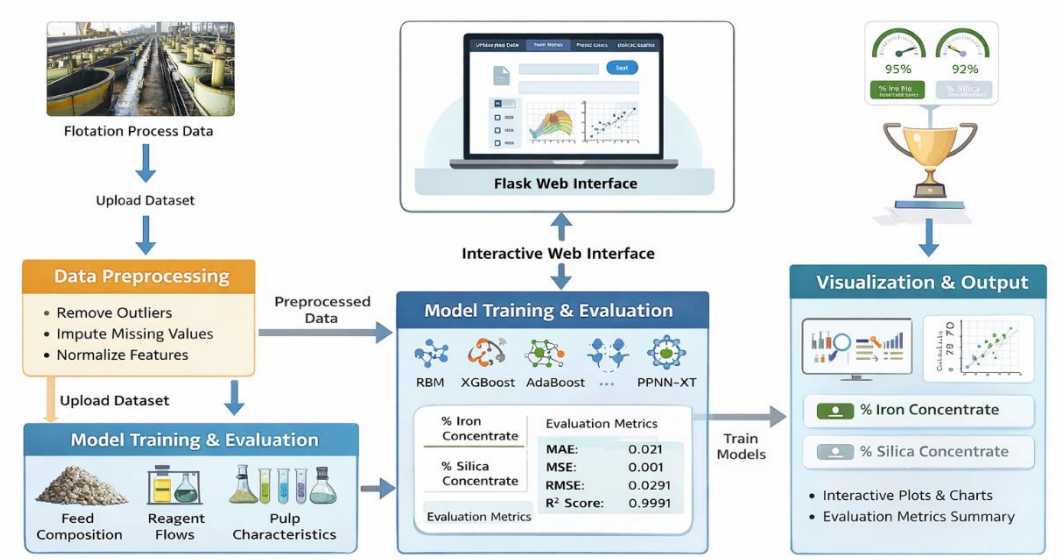

Figure 2: Proposed system architecture of ore quality prediction

1. Data Collection and Preprocessing: The system collects flotation process data from the dataset, including input features related to feed composition and operational parameters. The data is cleaned by removing irrelevant columns, handling missing values, and converting all features into numeric format. This ensures consistency and reliability for model training.

2. Feature Selection and Data Splitting: Relevant input features are selected by excluding target variables such as iron and silica concentrate values. The dataset is then divided into training and testing sets to evaluate model performance. This step ensures unbiased learning and proper validation of models.

3. Data Scaling and Normalization: The input features are scaled using standardization techniques to bring all variables to a common range. For specific models like RBM, additional normalization is applied to improve learning efficiency. This step enhances model stability and prediction accuracy.

4. Model Training and Learning: Multiple machine learning models (RBM, XGB, ADA, and PPNN-XT) are trained using the processed data. Each model learns patterns and relationships between input features and target outputs. The trained models are saved for future predictions to avoid repeated training.

5. Model Evaluation and Comparison: The performance of each model is evaluated using metrics such as MAE, MSE, RMSE, and R^2 score. Scatter plots are generated to compare actual and predicted values. This helps in identifying the best-performing model for accurate predictions.

6. Prediction (Single and Batch Mode): The system allows users to input data manually for single prediction or upload a dataset for batch prediction. The selected model processes the input and generates predictions for iron and silica concentrate values. This provides flexibility and real-time usability

7. Visualization and Decision Support: The predicted results are displayed along with graphical visualizations for better interpretation. These insights assist users in making informed decisions to optimize flotation process parameters. The system thus acts as a decision-support tool for industrial applications.

4. RESULTS ANALYSIS

The results section presents the key findings obtained from the study in a clear and structured manner. It highlights the outcomes of the analysis, showing how the data supports or contradicts the research objectives. Important patterns, trends, or relationships observed during the study are summarized to give a concise understanding of the findings. The section may include numerical data, comparisons, or visual representations to enhance clarity. It provides evidence-based insights that form the foundation for discussion and conclusions in the research.

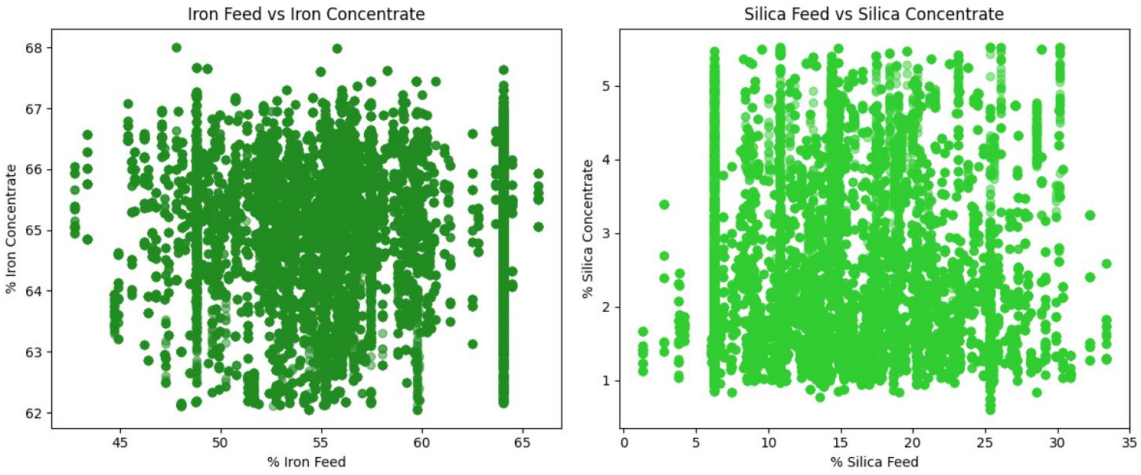

Figure 3: Feature Relationship Analysis: Feed vs Concentrate

The figure 3 illustrates the relationship between feed composition and concentrate quality for both iron and silica. The Iron Feed vs Iron Concentrate plot shows a weak to moderate positive relationship, indicating that higher iron feed generally contributes to higher iron concentrate, though with noticeable variability due to process conditions. In contrast, the Silica Feed vs Silica Concentrate plot shows a more dispersed pattern, suggesting a less direct and more complex relationship influenced by flotation parameters. These visualizations highlight the nonlinear nature of the process and justify the use of advanced machine learning models for accurate prediction.

Figure 4 presents scatter plot that demonstrates the comparison between actual and predicted values of % Iron Concentrate using the proposed PPNN-XT model. The data points lie almost perfectly along the ideal diagonal line, indicating extremely high prediction accuracy and minimal error. This strong alignment shows that the model effectively captures the underlying relationships in the dataset and produces highly consistent predictions. The near-perfect fit highlights the superior performance of the hybrid PPNN-XT model compared to other models, making it the most reliable approach for predicting concentrate quality in this system

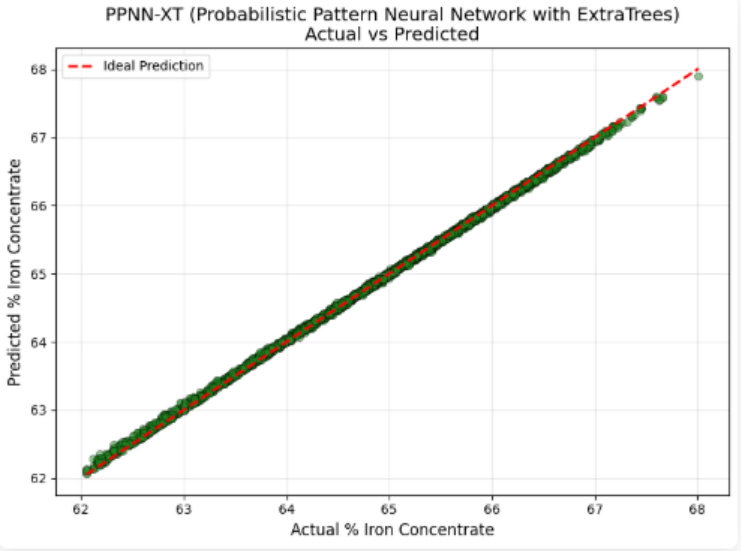

Figure 4: Illustration of scatter plot using PPNN-XT for iron concentrate%

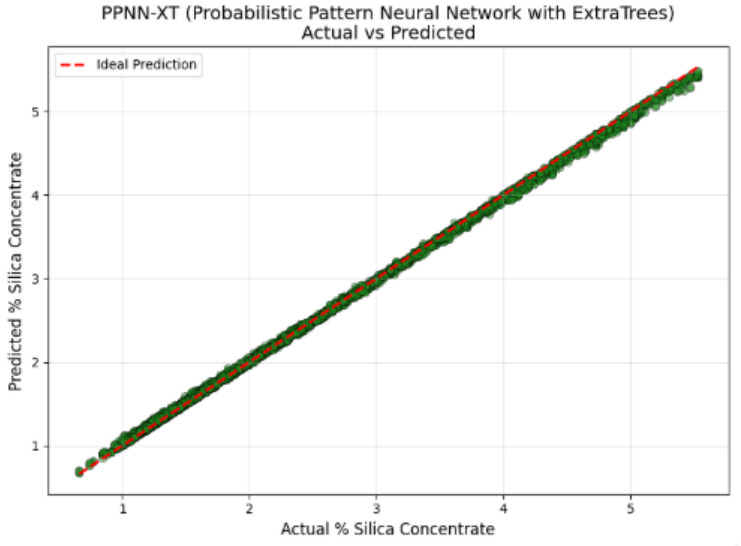

Figure 5: Illustration of scatter plot using Proposed PPNN-XT for silica concentrate%

Figure 5 shows scatter plot shows the comparison between actual and predicted values of % Silica Concentrate using the proposed PPNN-XT model. The data points are almost perfectly aligned along the ideal diagonal line, indicating extremely high prediction accuracy with minimal deviation. This strong alignment demonstrates that the model effectively captures the complex relationships within the dataset and produces highly consistent results. The near-perfect fit confirms that PPNN-XT significantly outperforms other models and is highly reliable for predicting silica concentrate quality

The figure 6 presents the final prediction results generated using the proposed PPNN-XT model for both % Iron Concentrate and % Silica Concentrate. Each sample corresponds to a set of input flotation parameters, and the model outputs predicted concentrate values with high consistency and precision. The results show stable iron concentrate predictions within an optimal range, while silica values remain relatively low, indicating effective impurity separation. This demonstrates the strong predictive capability and reliability of the PPNN-XT model in estimating concentrate quality, making it suitable for real-time decision support in flotation systems.

Sample	% Iron Concentrate (Predicted)	% Silica Concentrate (Predicted)
1	64.31	3.16
2	65.55	1.79
3	65.38	2.19
4	66.83	1.41
5	65.16	2.87
6	66.94	1.21
7	65.21	2.12
8	63.52	3.84
9	65.16	2.71
10	62.86	4.90

Figure 6: Prediction obtained using proposed PPNN-XT

4.1 Comparative analysis

The table 1 presents a comparative performance analysis of different regression models such as RBM, XGB Regressor, AdaBoost Regressor, and the proposed PPNN-XT model based on evaluation metrics such as MAE, MSE, RMSE, and R^2 Score for predicting concentrate quality. The RBM model shows relatively poor performance with high error values and a very low R^2 score (0.0088), indicating its inability to capture data variability. AdaBoost also exhibits moderate performance with higher errors and a low R^2 value (0.123), suggesting limited predictive capability. In contrast, the XGB Regressor demonstrates significantly better results with reduced error metrics and an improved R^2 score (0.5296), indicating a stronger ability to model nonlinear relationships. However, the proposed PPNN-XT model clearly outperforms all other models, achieving extremely low error values (MAE: 0.022, MSE: 0.0009, RMSE: 0.0304) and an almost perfect R^2 score (0.9992), which indicates near-ideal prediction accuracy. This highlights the effectiveness of the hybrid approach in capturing complex patterns and producing highly reliable predictions for iron and silica concentrate quality.

Table 1: Performance comparison of all the models for iron concentrate%

Model	MAE	MSE	RMSE	R^2 Score
RBM (Restricted Boltzmann Machine)	0.8682	1.2006	1.0957	0.0088
XGBoost Regressor	0.5798	0.5698	0.7548	0.5296
AdaBoost Regressor	0.8366	1.0622	1.0306	0.123
PPNN-XT (Probabilistic Pattern Neural Network with ExtraTrees)	0.022	0.0009	0.0304	0.9992

The table 2 presents the performance comparison of regression models such as RBM, XGB Regressor, AdaBoost Regressor, and the proposed PPNN-XT model for predicting % Silica Concentrate, evaluated using MAE, MSE, RMSE, and R^2 Score. The RBM model shows poor performance with high error values and a very low R^2 score (0.0189), indicating weak predictive capability. Similarly, the AdaBoost model also performs inadequately with relatively high errors and a low R^2 value (0.1132), suggesting limited ability to capture the complex relationships in silica data. The XGB Regressor performs significantly better, achieving lower error values and a moderate R^2 score (0.5559), demonstrating its effectiveness in modeling nonlinear patterns. However, the proposed PPNN-XT model clearly outperforms all other models, achieving extremely low error values (MAE: 0.0222, MSE: 0.001, RMSE: 0.0315) and an almost perfect R^2 score (0.9992), indicating highly accurate and reliable predictions. This confirms that the hybrid PPNN-XT model is the most effective approach for predicting silica concentrate quality.

Table 2: Performance comparison of all models for silica concentrate%

Model	MAE	MSE	RMSE	R^2 Score
RBM (Restricted Boltzmann Machine)	0.8824	1.196	1.0936	0.0189
XGBoost Regressor	0.5503	0.5413	0.7357	0.5559
AdaBoost Regressor	0.8864	1.081	1.0397	0.1132
PPNN-XT (Probabilistic Pattern Neural Network with ExtraTrees)	0.0222	0.001	0.0315	0.9992

5. CONCLUSION

The research demonstrates an intelligent and robust system for predicting % Iron Concentrate and % Silica Concentrate in flotation processes using advanced machine learning techniques. By leveraging models such as RBM, XGB, AdaBoost, and the proposed PPNN-XT, the system effectively analyzed complex relationships between flotation parameters and concentrate quality. Among all models, the proposed PPNN-XT demonstrated superior performance with minimal error and near-perfect accuracy, highlighting its ability to capture intricate data patterns and provide highly reliable predictions. The integration of data preprocessing, feature scaling, multi-model training, and visualization ensured a comprehensive analytical framework. Additionally, the implementation of a Flask-based web interface enabled user-friendly interaction for dataset upload, model training, prediction, and result visualization. The system provides a powerful decision-support tool for optimizing flotation operations, improving mineral recovery, and reducing impurities, thereby contributing to enhanced efficiency and productivity in mining processes.

REFERENCES

[1]. Ren, C.; Yang, J.; Liang, C. Estimation of copper concentrate grade based on color features and least-squares support vector regression. Physicochem. Probl. Miner. Process. 2015, 51, 163–172.

[2]. Monyake, K.; Han, T.; Ali, D.; Alagha, L.; Kumar, A. Experimental and Machine Learning Studies on Chitosan-Polyacrylamide Copolymers for Selective Separation of Metal Sulfides in the Froth Flotation Process. *Colloids Interfaces* **2023**, *7*, 41.

[3]. Allahkarami, E.; Nuri, O.S.; Abdollahzadeh, A.; Rezai, B.; Chegini, M. Estimation of Copper and Molybdenum Grades and Recoveries in the Industrial Flotation Plant Using the Artificial Neural Network. Int. J. Nonferrous Met. 2016, 5, 23–32.

[4]. H. Xie, Z. Mao, D. Xiao, and J. Liu, "Rapid detection of copper ore grade based on visible-infrared spectroscopy and TSVD-IVTELM," *Measurement*, vol. 203, Nov. 2022, Art. no. 112003, doi: 10.1016/j.measurement.2022.112003.

[5]. H. Yi and F. Li, "Assessment of grade fluctuation and quality characteristics of imported iron ore using neural networks," *Metal Mine.*, vol. 11, pp. 121–124, Nov. 2010.

[6]. A. K. Patel, S. Chatterjee, and A. K. Gorai, "Development of a machine vision system using the support vector machine regression (SVR) algorithm for the online prediction of iron ore grades," *Earth Sci. Informat.*, vol. 12, no. 2, pp. 197–210, Jun. 2019, doi: 10.1007/s12145-018-0370-6.

[7]. D. Xiao, C. Liu, and B. T. Le, "Detection method of TFe content of iron ore based on visible-infrared spectroscopy and IPSO-TELM neural network," *Infr. Phys. Technol.*, vol. 97, pp. 341–348, Mar. 2019, doi: 10.1016/j.infrared.2019.01.005.

[8]. Q. He, D. Wang, S. Liu, Y. Mao, H. Sun, and M. Sun, "in situ determination of BIF iron ore based on visible-near infrared spectral features," *Metal Mine.*, no. 12, pp. 140–145, Dec. 2018, doi: 10.19614/j.cnki.jsks.201812026.

[9]. M. Xu, Y. Mao, M. Zhang, D. Xiao, and H. Xie, "Rapid detection of iron ore and mining areas based on MSSA-BNVTELM, visible–infrared spectroscopy, and remote sensing," *Remote Sens.*, vol. 15, no. 16, p. 4100, Aug. 2023, doi: 10.3390/rs15164100.

[10]. Cook, R.; Monyake, K.C.; Hayat, M.B.; Kumar, A.; Alagha, L. Prediction of flotation efficiency of metal sulfides using an original hybrid machine learning model. Eng. Rep. 2020, 2, e12167.

[11]. Pu, Y.; Szmigiel, A.; Apel, D.B. Purities prediction in a manufacturing froth flotation plant: The deep learning techniques. Neural Comput. Appl. 2020, 32, 13639–13649.

[12]. Pu, Y.; Szmigiel, A.; Chen, J.; Apel, D.B. FlotationNet: A hierarchical deep learning network for froth flotation recovery prediction. Powder Technol. 2020, 375, 317–326.

[13]. Pural, Y.E. Developing a data-driven soft sensor to predict silicate impurity in iron ore flotation concentrate. Probl. Miner. Process. 2023, 59, 1–9.

[14]. Ali, D.; Hayat, M.B.; Alagha, L.; Molatlhegi, O.K. An evaluation of machine learning and artificial intelligence models for predicting the flotation behavior of fine high-ash coal. Adv. Powder Technol. 2018, 29, 3493–3506.

[15]. Gomez-Flores, A.; Heyes, G.W.; Ilyas, S.; Kim, H. Prediction of grade and recovery in flotation from physicochemical and operational aspects using machine learning models. Miner. Eng. 2022, 183, 107627.